

2017-03-28

Aiding Soundtrack Composer Creativity through Automated Film Script-profiled Algorithmic Composition

Kirke, Alexis

<http://hdl.handle.net/10026.1/12718>

10.5920/JCMS.2017.01

Journal of Creative Music Systems

University of Huddersfield Press

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

Aiding Soundtrack Composer Creativity through Automated Film Script-profiled Algorithmic Composition

Alexis Kirke, Eduardo Miranda

[Article details](#)
[Body of article](#)
[References](#)
[Footnotes](#)
[Author contacts](#)
[Issue index](#)

Keywords

Movie script, soundtrack, music and emotion, algorithmic composition

Abstract

The creation of soundtracks for feature length films is a major undertaking. One method involves watching the film for sections that inspire some emotional response, and then improvising live. The composing of a score is often a combination of bottom up and top down processes. For example small segments of score are composed first, while in parallel a broader structure is planned based on the narrative. We present a system that attempts to aid the composer in this creative process. The system focuses on the script rather than moving image of the film. This has two advantages: (a) a composer can begin their work earlier in the process of film production; (b) the automated analysis of text is simpler than the automated analysis of video and speech. The system performs an automated analysis that approximates the script structure. It also attempts, using word analysis, to give insight into the emotional trajectory of the story and the characters. These are then used as part of an algorithmic creativity system that suggests musical sketches to the composer for parts of the script.

1 Introduction

There is a traditional process used for creating a film soundtrack (Jewell, 2007): screening, spotting, laying out, motif development and final composition. The screening involves the composer viewing a cut to provide them with an initial sense of the film. Spotting involves deciding which parts of the film require music, and the type of music required. Laying out involves converting these decisions to a score-based label sequence indicating visually where music is needed. Musical sections in films are typically short, thus allowing the director to re-cut with greater ease. This is supported by many composers using short motifs – often motifs for characters (Green, 2010) – and film themes.

The average length of a high-grossing feature film is almost two hours (Galician & Bourdeau, 2008). Even if only fifty percent of the film contains a scored soundtrack, this requires an hour of music; music which must be fully scored and recorded. There are many examples where composers find it helpful to begin this process before the film is available (Karlin & Wright, 2004). This requires them to work from a script. Reading and absorbing scripts is a time consuming process. Developing an understanding of the script structure is an additional strain.

In a bid to address some of the time pressures of modern film production (Tanrisever, 2001; Kulezic-Wilson, 2015; Bettinson, 2013), a system called *TRAC* (Textual Research for Affective Composition) has been developed. *TRAC* attempts to parse the script and highlight structural and affective elements. Tools are provided to generate short soundtrack examples based on these affective elements. The process is very approximate, because narrative has a highly complex cognitive and affective impact. However, for a composer approaching a script who is under time pressure, it can provide a valuable “in” for kick-starting and guiding some of their work process. Even ignoring such time constraints on a sound-track composer, the system is still potentially helpful from the perspective of structure and organisation.

2 Related Work

The core approach of *TRAC* is to auto-analyse a script to seek emotional words and use these to generate affective musical features for a sound-track. Thus the related work section will look at three elements: affective music composition, script analysis, and assisted sound-tracking .

2.1 Affective Music Composition

TRAC utilises an affective music composition algorithm based on a dimensional model of emotion. Investigations into the affective impact of audio-visual media have long been a part of researching the psychology of emotions (Sloboda, 1986; Seashore, 1938). A common area of investigation has been music, during which various musical features have been tested to examine the emotions they communicate or cause (Williams et al., 2014). For example: major key modes tend to communicate

positivity, and minor key modes often communicate negativity. In addition lower tempo music in a major key communicates relaxation, and high tempo music without a key (atonal) communicates fear or anger.

The two most common models of emotion used in studies are categorical and dimensional. Categorical models incorporate lists of basic emotions: "happiness", "sadness", etc. The dimensional models usually have a set of two or more continuous properties within which emotional states can be situated.

The most common dimensional model is the valence/arousal model. Valence is a measure of emotional positivity, and arousal a measure of physical activity. Thus happy is mapped to high arousal and high valence, and sad to low arousal and low valence. A third dimension, called dominance, is sometimes added to the valence/arousal model (Marsella et al., 2010). It can be thought of as the degree of personal power in the emotion. Dominance is most intuitively explained as follows. Consider that fear and anger can both be mapped to high arousal and low valence. However fear is also mapped to low dominance, and anger to high dominance. The valence/arousal/dominance model is used in TRAC.

There is a long history of computers being used to automatically compose. This is known as algorithmic composition (Miranda, 2001; Fernandez & Vico, 2013). In particular there are a number of computer composition systems like TRAC that generate music based on an affective profile. In these, the user can select an emotional trajectory or description and the system will attempt to auto-compose music based on that requirement (Kirke & Miranda, 2011; Huang & Lin, 2013). A fuller survey of such systems can be found in Williams et al. (2014). TRAC differs from most of these because it generates the emotional trajectory / description based on automated script analysis. Furthermore, most of these do not attempt to utilise the dominance dimension.

2.2 Script Analysis

Many of the systems described below are rooted in natural language processing and generation. An overview of this topic can be found in (Clark et al., 2010). TRAC utilises a standard set of such natural language tools (Bird et al., 2009). Clearly, character dialogue is one indicator of a script's affective content, and it is one element used in TRAC's analysis. Searching for emotional words spoken by characters in script dialogues is an approach used in (Park et al., 2012). A broader data driven analysis of the affective content of script stories can be found in (Bellegarda, 2010; Murtagh & Ganz, 2014) which apply a pattern recognition-based narrative analysis of emotion to the film *Casablanca* and the novel *Madame Bovary*.

As previously mentioned, TRAC attempts to generate music based on the script textual structure. However there are other systems that aim to generate new dramatic material through a deeper structural analysis. Munishkina et al. (2013) automatically analyses scripts from Internet Movie Script Database. Their system parses scenes, and scene locations, as well as dialogues and events. It uses TF-IDF clustering to gather these scenes together and then create levels for a computer game – in this case a *Raiders of the Lost Ark*-based game. Walker et al. (2011) is another system that aims to develop materials for a game. The system auto-generates character models using six film characters, including Indiana Jones from *Raiders of the Lost Ark*, and characters from *Pulp Fiction*. Another part of the system is used to re-produce the characters' dialogue but with the speaking style of a different character. The system was used to generate text for a role-playing game. Successful tests were run to examine if users could differentiate the characters based on speaking style. Much of this work was based on ideas and technology developed previously in Lin & Walker (2011).

There are further systems that – like the above – utilise a deeper structural analysis than TRAC, but have a more explicitly commercial purpose. For example, Sang & Xu (2010) perform some initial testing on a system for generating automated movie summaries. These summaries are produced for the purposes of movie marketing. The system cross-references the movie with its script, allowing a segmentation of the film. Then it attempts to track character interactions and sub-plots. Eliashberg et al. (2007) aims to predict movie hits. This involves incorporating a summarization system into a broader statistical learning model. Natural language processing and extracted screenwriting domain knowledge are part of this approach for predicting a movie's profit margin. Kundu et al. (2013) also uses natural language processing, combining it with genetic algorithm techniques, to segment scripts into scenes. The approach reaches an accuracy of 45% across three films.

ScriptThreads (Hoyt et al., 2014) has some similarities to TRAC. It parses screenplays, generating visualisations to aid the understanding of character behaviour. These visualisations indicate how characters interact with each other. This is demonstrated using the feature film *The Big Sleep*.

2.3 Assisted Sound Tracking

It is well known that music can change the emotional impact of a film (Boltz, 2004). Physiological studies on how music affects people's viewing show that the results are not necessarily predictable (Ellis, 2005). However, patterns have been found. Blumstein et al. (2010) found some consistency in the ways that sound is used by filmmakers to manipulate the emotions of the audience. Cohen (2013) develops a framework for understanding how music tracks work from the perspective of narrative.

Although TRAC aims to generate new music "from scratch", there have been a number of related systems that focus on pre-created music. These systems aim to automatically match such music to given visual content. Picasso (Stupar, 2013) analyses a database of movies, developing a model linking movie screenshots and the composer music that is playing at that time the shots are onscreen. Suppose a user wishes to find music for a new photo. When Picasso is presented with the photo, its model is searched for the screenshot most similar to the photo. The music that was playing with this

screenshot is used as a template to search the user's own musical database. The most similar track from the user's database is suggested as a soundtrack for the image. Gomes et al. (2013) builds a similar model using a genetic algorithm approach. Musical features (for example loudness) are mapped to various visual features (for example the amount of movement onscreen). Thus the system attempts to find the optimal piece of music for a specific piece of video. Kuo et al. (2013) builds a model by applying Latent Semantic Analysis to YouTube videos. Yin et al. (2014) is a system focusing on home movie soundtrack selection. The user is required to manually tag video features. This tagging is combined with a number of automated methods. These methods examine pitch tempo patterns and motion-direction.

The above systems select music. There are a number of systems that generate the music themselves. For example: CBS (Jewell et al., 2003). CBS is a concept system that involves the manual markup of the movies, providing an object-based overview. These objects give indications of mood, characters in the story and events in the narrative. The music generation utilises genetic algorithms (*TRAC* uses a rule-based rather than evolutionary algorithm) to match the composition parameters in this mark-up. They illustrate this with an example: "a 'happy' scene may be defined to use a major pitch mesh with violins and flutes providing a light texture, whereas an 'unhappy' scene may use a minor pitch mesh with 'celli and clarinets". RaPScoM (Doppler et al., 2011) attempts to automate the process of semantic mark up of the movie. Features include Action/Violence, Fear/Tension, Freedom, Joy/Comedy, etc. The algorithmic composition system – a hierarchical approach based on Markov models of melodies – generates melodies and a simple accompaniment. Alternatively it can transform user-provided MIDI material. It is not clear whether RaPScoM has been fully implemented. *TRAC* also uses a hierarchical generative approach.

Vane & Cowan (2007) and Hedemann et al. (2008) both recognise the advantages of providing specialised software for sound-track composers. They incorporate elements such as a repetition tool, and a tool for fitting sequences within specified time boundaries. Davis et al. (2014) involves a system that – like *TRAC* – generates music from text. The system is called *TransPose* and focuses on novels rather than film scripts. Like *TRAC*, *TransPose* utilises an emotional word search.

Based on the related work discussed in Section 2, the originality of *TRAC* stems from a combination of the following elements: (i) it has been implemented; (ii) it does not involve manual mark-up; (iii) it generates rather than selects music; and (iv) it utilises a three-dimensional model of emotion.

3 System Description

For the purposes of explanation *TRAC* will be broadly divided into two parts: Script Analysis and Music Sketching.

3.1 Script Analysis

The script analysis consists of three elements. Firstly a text file containing the script is imported. The text is stripped of extraneous spaces and carriage returns. Some scripts contain additional text that needs to be stripped. For example "Page 1" or "Continued:". Although formal scripts have a standard format, there are usually small discrepancies in those available on the internet. Fig. 1 shows four of the most common types formats encountered. The top left script is formatted in the formal idiom.

Next the script is broken down into objects for analysis. At this point additional technical differences must be addressed to enable identification of objects. Scene headings may use "INT." or "INTERIOR:". Dialogue may be formatted with the character's name centered over the text, or with the character's name in the same paragraph as the dialogue.

The content of a movie script can be divided into three types of information: scenes, action and dialogue. Scenes are (usually single line) headings that describe where events in the following text are occurring. A scene heading will typically indicate: whether actions occur indoors or outdoors (INT or EXT), the broad location of the action, and sometimes a more detailed location description. For example: "EXT. – TATOOINE – WASTELAND". Everything following this line up to the next scene heading will occur in the same location. Dialogue information is indicated by a character's name in upper case, followed by a continuous paragraph of text starting on the next line. Action information looks similar to dialogue, but without a character name headline. Action information describes what happens in the story when dialogue is not sufficiently clear or relevant.

There are additional, though less common, standardised elements in scripts. From the point of view of the analysis being done here, they are incorporated into one of the above three objects. An example of this is when a piece of dialogue words in parentheses between the character name headline and the spoken text. The parentheses are a modifier, implying mood or emphasis. In *TRAC* they are incorporated into the dialogue object as another indicator of affective state. The precise rules used by *TRAC* for parsing scripts are not detailed here. The rules were developed through a manual analysis of the four types of scripts shown in Fig. 1.

<p>Pursued by the Empire's sinister agents, Princess Leia races home aboard her starship, custodian of the stolen plans that can save her people and restore freedom to the galaxy...</p> <p>The awesome yellow planet of Tatooine emerges from a total eclipse. Her two moons glowing against the darkness, a tiny silver spacecraft, a Rebel Blockade Runner firing lasers from the back of the ship, races through space. It is pursued by a giant Imperial stardestroyer. Hundreds of deadly laserbolts streak from the Imperial Stardestroyer, causing the main solar fin of the Rebel craft to disintegrate.</p> <p>INT. REBEL BLOCKADE RUNNER - MAIN PASSAGEWAY</p> <p>An anti-air robot, the ship's two robots, Artoo (droid 02-00) and Chewbacca (C-3PO) struggle to make their way through the shaking, bouncing passageway. Both robots are old and battered. Artoo is a short, close-armed tripod. His face is a mix of computer lights surrounding a radar eye. Chewbacca, on the other hand, is a tall, slender robot of human proportions. He has a gleaming bronze-like metallic surface of an Art Deco design.</p> <p>Another blast shakes them as they struggle along their way.</p> <p>CHWEEBACCA Did you hear that? They've shut down the main reactor. We'll be destroyed for sure. This is madness!</p>	<p>Her pg. 3</p> <p>INT. THEODORE'S OFFICE - EARLY EVENING</p> <p>Theodore walks through the reception area. The office is almost empty except for him and the receptionist, PAUL. Theodore begins to scan each letter through a scanner on the front desk, then puts them in the outgoing mail box.</p> <p>Paul is sitting at a desk across the room, reading handwritten letters on a computer monitor.</p> <p>PAUL Theodore! Letter writer 612.</p> <p>THEODORE Hey, Paul.</p>
---	---

Rebel troopers rush past the robots and take up positions in the main passageway. They aim their weapons toward the door.

THREEPIO
We're doomed!

The little R2 unit makes a series of electronic sounds that only another robot could understand.

THREEPIO
There'll be no escape for the Princess this time.

Artoo continues making beeping sounds. Tension mounts as loud metallic latches clank and the scream of heavy equipment are heard moving around the outside hull of the ship.

THREEPIO
What's that?

EXT. SPACECRAFT IN SPACE
The Imperial craft has easily overtaken the Rebel blockade

1 BLACKNESS
Then two faint lights appear, close together... growing brighter. They resolve into two DEEP SUBMERSIBLES, free-falling toward us like express elevators.

One is ahead of the other, and passes close enough to FILL FRAME, looking like a spacecraft blazing with lights, bristling with insectile manipulators.

TILTING DOWN to follow it as it descends away into the limitless blackness below. Soon they are fireflies, then stars. Then gone.

CUT TO:
PAN DOWN to reveal a small space cruiser heading TOWARD CAMERA at great speed. JAB with the cruiser as it heads toward the beautiful green planet of Naboo, which is surrounded by hundreds of Trade Federation battleships.

INT. REPUBLIC CRUISER - COCKPIT
In the cockpit of the cruise, the CAPTAIN and PILOT maneuver closer to one of the battleships.

QUI-GON : (off screen voice) Captain.

The Captain turns to an unseen figure sitting behind her.

CAPTAIN : Yes, sir?

QUI-GON : (V.O) Tell them we wish to board at once.

CAPTAIN : Yes, sir.

The CAPTAIN looks to her view screen, where NUTE GUNRAY, a Neimoidian trade viceroy, waits for a reply.

CAPTAIN : (cont'd) With all due respect for the Trade Federation, the Ambassadors for the Supreme Chancellor wish to board immediately.

NUTE : Yes, yes, of course... Ahh... as you know, our blockade is perfectly legal, and we'd be happy to receive the Ambassador... happy to.

The screen goes black. Out the cockpit window, the sinister battleship looms ever closer.

3 EXT. THE BOTTOM OF THE SEA
On the other side, craned into the remaining space is a bearded wide-body named LEWIS BODINE, who is also asleep. Lewis is an R.O.V. (REMOTELY OPERATED VEHICLE) pilot and is the resident Titanic expert.

Anatoly glances at the bottom sonar and makes a ballast adjustment.

CUT TO:
The screen goes black. Out the cockpit window, the sinister battleship looms ever closer.

Figure 1 – Four examples of script formats available online. The script in the top left of the figure is a formal screenplay.

Not all script files have page numbers, and not all script fonts are the same. Thus a generated page size is used – the virtual page size. Virtual pages in *TRAC* are significantly larger than standard script page sizes. This has the advantage of increasing the sample size of the words used to calculate affective results.

The results of the script analysis can be broadly divided into two types: structural and emotional. The simplest structural measure involves estimating who the main characters are in the script. This is done by estimating each character's activity: the letter count of their entire dialogue, added to the letter count of all actions in which their name is mentioned. The characters' activity can be displayed in decreasing order for the user. Another structural measure in *TRAC* is the activity trajectory of a given single character: for example the protagonist. This is a plot of the character's total activity on each virtual page (vPage). A plot is also provided of the amount of dialogue on each vPage in the script. The identification of sections of script without dialogue is useful as it often indicates a need for significant sound-tracking. Additional plots can be requested on a character-by-character basis of a measure called action domination. For each vPage, action domination is defined as the letter count for actions involving a character, divided by the number of letters of their dialogue.

There is another indicator of visual activity: the number of scenes per page, called here scene rapidity. If a character moves rapidly from place to place, this may lead to the script moving rapidly from scene to scene. Even if the character is not moving rapidly from place to place, such scene shifts will often be done to increase tension or a sense of activity. Conversely, a low scene rate highlights the location of a longer scene. Such longer scenes can help the composer to “get their teeth into” the scoring. This is one way that film composers approach scoring: find longer sections of the movie that have a significant dramatic or visual arc. Such sections provide potential opportunities to write complete pieces of music.

The other main approach to the script analysis is affective analysis. The general problem of text sentiment discernment has no reliable and robust automated solution. In *TRAC* such an analysis is approximated using word-by-word sentiment. A word-approach can only indicate sentiment because words modify each other. For example the phrase “this day is bad” and “this day is the opposite of bad” have opposing affective implications. The phrase “the opposite of” reverses the sentiment of the word “bad”. However, word-by-word sentiment analysis is a good first approximation, particularly in the context of enabling, as opposed to replacing, the composer.

The *TRAC* sentiment analysis utilises the ANEW database (Bradley & Lang, 1999) and the Warriner et al. (2013) database. ANEW and Warriner et al. (2013) are databases that provide estimates of valence, arousal and dominance for many thousands of English words. Before applying these databases to the affective analysis of a piece of script text, the words in the text are stemmed and stop-words (Bird et al., 2009) are removed. Stemming standardises words by removing common additions. For example removing “ing” from “adding” and “ly” from “seriously”. Stop words are common words like “and”, “or” and “the”. The *Python NLP* toolbox (Bird et al., 2009) is used for both stemming and stop word removal.

The resulting text is then broken down into individual words. If any of these words are found in the ANEW or Warriner et al. (2013) database, then their valence, arousal and dominance are noted and normalised between 0 and 1. All normalised values for a piece of text are then averaged across that piece of text, giving an estimate of the affective content for that text. Thus valence, arousal and dominance trajectories are provided across vPages, on a character by character basis. Similarly the valence, arousal and dominance of a whole vPage across all characters can be calculated, labelled here as the global affective trajectory.

3.2 Music Sketching

The *TRAC* music sketching tool utilises the calculations of valence, arousal and dominance from the script analysis. However it is not the raw valence, arousal and dominance calculated in Section 3.1 that are used in the composition. The values are scaled first, using the mean valence, arousal and dominance across the whole script. Reasons for doing this include the fact that the mean affectivity

parameter values based on word analysis were found to be greater than 0.4 in the examples we tested. Thus the values must be shifted downwards to take this bias into account. Otherwise the entire affective nature of the music would tend to focus on the upper two thirds of the valence, arousal and dominance space.

Algorithmic composition systems designed to communicate valence and arousal most commonly implement the relationships shown in the first four rows of Tab. 1. These four rows and the next three rows are simplified versions of the Livingstone et al. (2010) affective composition rules. The next three rows (low valence / high arousal) utilise atonal music to communicate the affective states (Kirke & Miranda, 2015). The final two rows relate to low and high dominance, fear and anger respectively. Le Groux & Verschure (2012) supports the approach of using low tempo to communicate low dominance, and also introduce tentative results that low pitch communicates a lower dominance than a high pitch. However Tab. 1 is a simplification of their results, as they also indicate that a medium pitch can actually communicate a lower dominance than a low pitch. The rule in Tab. 1 simplifies their results in a best fit linear trend.

The rules in Tab. 1 are implemented as transformations rules on MIDI files. The specific transformations are shown in Tab. 2. These transformations act on source material which can be provided by the composer, or generated by *TRAC*. The focus here will be on *TRAC*-generated phrases. The phrase generation system is based on a random walk around middle C. At each time step the pitch can go up or down by one or two halftones with equal probability, and six halftones with a lower probability. At each time step there is also a chance at each step of jumping back to middle C, and a smaller chance of jumping up to G below middle C, in addition to the walk probabilities. This is to help establish the keys of C Major or C Minor. Such an algorithm could be viewed as overly simplistic. However it was found that most of the affective expressiveness generated by *TRAC* comes from the transformations performed on these simple tunes.

Affectivity	Music Features
High Valence	Key Mode Major
Low Valence	Key Mode Minor
High Arousal	Increase Tempo
Low Arousal	Decrease Tempo
High Valence & Arousal	Also increase pitch
Low Valence & Arousal	Also decrease pitch
High Valence & Low Arousal	Also increase pitch
Low Valence & High Arousal	Atonal
Low Valence & High Arousal & Low Dominance	Also decrease pitch and tempo
Low Valence & High Arousal & High Dominance	Also increase pitch

Table 1 – Affectivity Rules.

Music or Affective Features	Note Transform
Key Mode Major	Move pitch to nearest C major note.
Key Mode Minor	Move pitch to nearest C minor note.
Low Tempo	Tempo = arousal * 150
Low Tempo	Tempo = arousal * 150
High pitch	Add to MIDI Pitch 48*valence-24
Low pitch	Add to MIDI Pitch 48*valence-24
High pitch	Add to MIDI Pitch 48*valence-24
Atonal	Randomly transpose by a halfnote.
Low Valence & High Arousal & Low Dominance	Add to MIDI Pitch 48*dominance-24 Tempo = dominance * 150
Low Valence & High Arousal & High Dominance	Add to MIDI Pitch 48*dominance-24

Table 2 – Affectivity transformations applied to MIDI values.

The randomly generated phrases are combined to create themes. A theme is generated from a phrase by randomly selecting one of five strategies: (a) repeat the phrase (two to four repetitions, selected randomly); (b) add the retrograde of the phrase to the end of the phrase; (c) repeat the following transformation from one to three times: randomly adjust the pitches of two notes in the phrase (within the boundaries of the maximum / minimum pitches currently found in the phrase), concatenate the results to the end of the phrase; (d) add one to three random shuffles of the notes at the end of the phrase; (e) add one to three “rotate and adjust” instances to the end of the phrase – each instance involves rotating the phrase across the time axis in the increasing time direction, (allowing wrap-around), at each rotation randomly adjusting one note’s pitch.

Themes are combined to create sections, constrained by a user-specified pattern. The user can specify structures such ABA or ABCA. A, B, C and D are generic theme labels. “A” always refers to the pre-provided theme, either generated randomly by *TRAC*, or provided by the user. The other themes (B, C, D, etc.) will be generated by *TRAC* before building the section. The user communicates the structures to *TRAC* in a numerical form i.e., 010 for ABA, 0120 for ABCD, and so forth.

The section is then transformed based on valence, arousal and dominance estimates from the desired section of the script. For example, suppose the user asks *TRAC* to randomly generate a phrase, say A, for the protagonist. Then the user may decide they want to hear some sketches using the protagonist theme for vPages 12 and 25. In this case they specify that the musical structure for

vPage 12 is 010 (ABA) and for vPage 25 is 0120 (ABCA). The user then instructs *TRAC* to generate the two transformed sections. These two sketches will be based on the valence / arousal / dominance measurements on vPage 12 and vPage 25 respectively. The user listens to the results and can trigger *TRAC* to re-compose should they wish to hear more examples.

4 System Examples

The scripts of four well-known films will be utilised: *Alien* (Scott, 1979), *Star Wars* (Lucas, 1977), *The Matrix* (Wachowski et al., 1999) and *The Godfather* (Coppola, 1972). Examples will be given of the structural *TRAC* measures. This will be followed by examples of the affective *TRAC* measures.

4.1 Structural Measures

Structural measures provide indications of script features not directly related to affectivity. The Character Activity measure for some of the key characters in *Alien*, *Star Wars*, *The Matrix* and *The Godfather* are shown below. They are in reverse order of the number of words spoken by the character:

[[4, "DALLAS", 16526], [10, "RIPLEY", 13986], [3, "PARKER", 11179], [13, "ASH", 8788], [6, "KANE", 6024], [5, "LAMBERT", 5075],...]

[[6, "LUKE", 42467], [3, "BEN", 18741], [38, "HAN", 15101], [27, "THREEPIO", 13773], [63, "VADER", 6449], [68, "LEIA", 5092],...]

[[28, "MORPHEUS", 17143], [32, "NEO", 17083], [11, "TRINITY", 11804], [12, "CYPHER", 5752], [13, "TANK", 5283], [35, "AGENT SMITH", 5248],...]

[[28, "MICHAEL", 27926], [2, "SONNY", 15638], [37, "HAGEN", 11726], [51, "DON CORLEONE", 9803], [33, "CLEMENZA", 6712],...]

The protagonists of these films are known to be RIPLEY, LUKE, NEO and MICHAEL respectively. The flaws of the Character Activity measure as a way of identifying key characters are evident in the list of names from *Alien*. The ALIEN does not occur in the list because it has no dialogue. Furthermore DALLAS' prominence over RIPLEY is because RIPLEY is less talkative, not because DALLAS is more important dramatically.

Fig. 2a, 2b, 2c and 2d plot the activity measures for the top three characters in *Star Wars* and *The Godfather*. Fig. 2a shows that in *Star Wars*, LUKE is across most of the script. HAN and BEN appear only for certain segments. In BEN's case this is due to his death. For HAN, it is because he is not in the film from the start, and takes his leave temporarily after the Death Star escape. Fig. 2b shows that in *The Godfather*, MICHAEL is inactive in the script for a significant part. This is when he is away in Sicily. Fig. 2b also captures the fact that SONNY dies midway through the movie. The activity and inactivity highlighted here are all significant story points.

Fig. 3 plots the Action Domination for the top three characters in *Star Wars* and *The Matrix*. For LUKE, Fig. 3a's highest peak is the section leading up to his destruction of the Death Star, as is the second highest peak. LUKE's third highest peak occurs during the rescue of Princess Leia and their swinging across a chasm. As is often the case in action scenes, such segments in *Star Wars* are heavily sound-tracked in the final movie. BEN's highest action domination peak occurs during his fight with VADER. HAN's highest peak occurs during his running battle in the corridors of the Death Star. NEO's highest peak in Fig. 3b occurs during his final fight and chase scene with the antagonist (SMITH). NEO's second highest peak occurs at the point he wakes up in the real world for the first time. This segment has no dialogue and is heavily sound-tracked in the final movie. TRINITY's action domination peak occurs during the opening action scene. MORPHEUS' peak occurs during his escape after being interrogated by SMITH – another dialogue-light action scene.

Fig. 4 shows the Scene Rapidity measure for the films *Alien* and *The Matrix*. The highest scene rate peak in *Alien* – Fig. 4a – is located at the film's climax (which has almost no dialogue and thus was heavily sound-track). The second highest peak occurs at the opening of the film. This opening is heavily sound-tracked, and involves a series of changing shots establishing the spaceship set. The third peak in Fig. 4a occurs when DALLAS is attacked by the ALIEN in the ship's vent system.

In Fig. 4b (*The Matrix*) the local maximum at vPage 2 occurs where TRINITY is running for her life. The next local maximum, at vPages 18 and 19, occurs when NEO awakens in the real world. The highest peak occurs at vPage 29. This peak is the build-up starting with NEO's training, leading to the betrayal by TANK, and thus the characters fleeing. The second highest peak in Fig. 4b occurs at the film's climax. NEO fights, flees and becomes "The One".

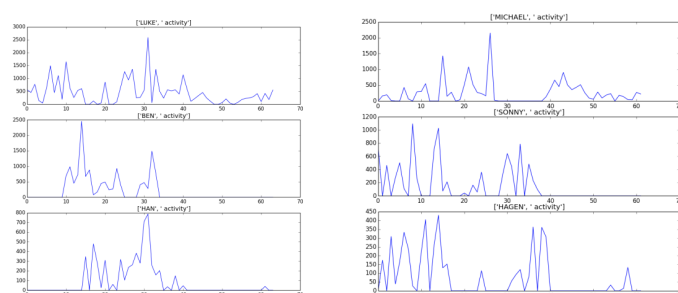


Figure 2a and 2b – Activity graphs for top 3 characters for *Star Wars* (a) and *The Godfather* (b).

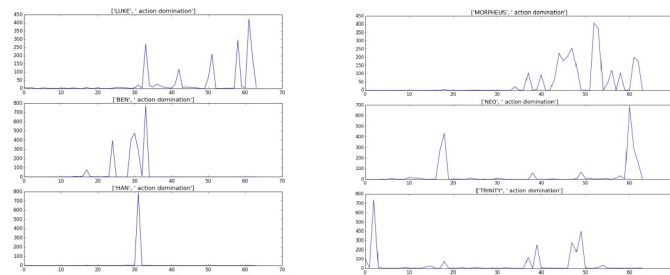


Figure 3a and 3b – Action Domination graphs for top 3 characters for *Star Wars* (a) and *The Matrix* (b).

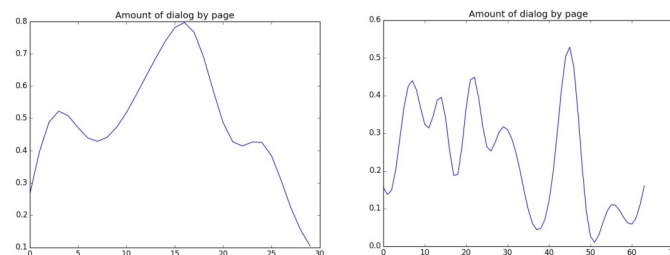


Figure 4a and 4b – Scene Rapidity and Dialogue for *Alien* (a) and *The Matrix* (b).

4.2 Affective Measures

The music sketching tool in *TRAC* is based on affective measures. Plots of these affective are shown in Fig. 5. This figure contains global affectivity plots for *Star Wars* and *The Godfather*. These graphs are less useful from a structural point of view. The affective analysis is necessarily an approximation. This is partly because it is word-based rather than phrase-based, and partly because non-word-based affective impact (for example changes in a character's behaviour) are not taken into account. However a few structural elements are captured by the plots in Fig. 5. Fig. 5a shows the global valence of *Star Wars* reaching a climax at the script's end. The local minima leading up to this climax are coincident with VADER killing LUKE's fellow X-Wing pilots (as the Death Star prepares to fire on Yavin). The largest minimum occurs during the destruction of an X-Wing and the first (missed) attempt to hit the Death Star's vulnerable vent. The other large local minimum occurs approximately half way through the *Star Wars* script. At this point, the heroes are trapped within the Death Star. They are forced to hide in a garbage compactor which almost kills them.

In Fig. 5b (*Alien*) the first valence local minimum occurs when the crew realise they've been woken up early and that they are not yet home. The final large valence minimum occurs when PARKER is nearly killed by depressurization, and when ASH attempts to kill RIPLEY. Note that this period is an arousal local maximum and valence local minimum, consistent with anger or fear. The dominance is low around vPage 25, which indicates the affective state of this period as fearful. This is appropriate because PARKER fears death and RIPLEY fears being killed by ASH. Global arousal moves upwards to a climax at the end of the script. The script end, in large part, consists of RIPLEY running and panicking.

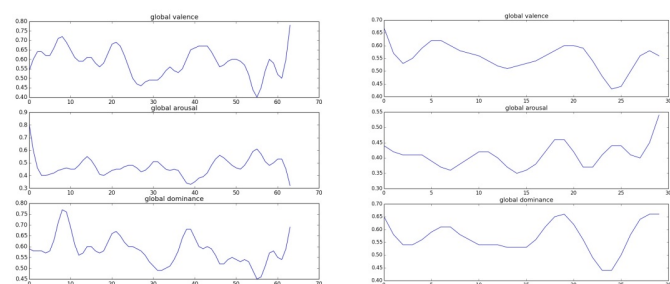


Figure 5a and 5b – Graphs of Global Affectivity for *Star Wars* (a) and *Alien* (b).

The bounds on the y-axis of the graphs in Fig. 5 provide some structural indications. The *Star Wars* script is more emotionally positive and aroused – in word terms – than *Alien*. This is consistent with their genres. *Alien* is a Sci-fi Horror with a sad ending. *Star Wars* is a Space Opera with a happy ending. These differences are audible in the sound-tracks of the final films.

In addition to global profiles, *TRAC* generates affective profiles for characters, one of the most important being the protagonist. Fig. 6 contains plots giving affective profiles for the protagonists of *The Godfather* (6a) and *Alien* (6b). The global minimum for MICHAEL'S valence coincides with the shooting of his father. The other two minima occur when MICHAEL discovers his father vulnerable at the hospital, and his wife is murdered.

It is worth repeating that these graphs are of limited structural use to the composer. Our overall emotional reaction to a film comes from identification with characters. This includes our imagination about their internal states. Such states can only be indicated by dialogue and actions. In addition the emotional impact of a story is cumulative. Whereas the affective text analysis here is not only independent of the words around a word, but also of the scenes around a scene.

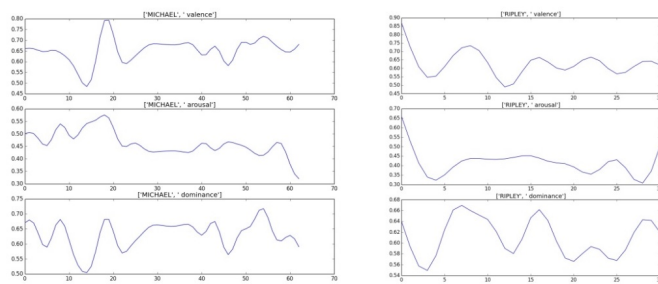


Figure 6a and 6b – Affective Profiles for Michael Corleone (a) and Ellen Ripley (b), the protagonists of *The Godfather* and *Alien* respectively.

TRAC was used to generate musical suggestions for the following script segments:

1. Global affect: the Alien revealing itself in a craft Ripley has used to escape (*Alien* vPage 29)
2. Global affect: the Millennium Falcon's escape from the Death Star (*Star Wars* vPage 38)
3. Neo's affect: towards the end of the film when he tries to escape from Agent Smith (*The Matrix* vPage 61)
4. Global affect: the opening monologue at a wedding about a daughter being attacked (*The Godfather* vPage 0)

For each of the four, the first suggestion was taken. The results are shown in Figs. 7–10 respectively. They can also be heard online (Kirke, 2016a–d). The scene where Ripley realises the alien is on board her escape craft is estimated as being valence 0.01, arousal 0.89 and dominance 0.07. The resulting music in Fig. 7 is low pitch, low tempo and atonal – which seems appropriate for the dark fearful and mysterious scene. The next scene, where the Millennium Falcon has escaped the Death Star (Fig. 2) is estimated as being valence 0.66, arousal 0.27 and dominance 0.72. The suggested music in Fig. 8 is medium-high pitch, medium tempo and major key mode. It could be interpreted as triumphal. *The Matrix* scene where Neo is fleeing Agent Smith is estimated as being valence 0.49, arousal 0.57 and dominance 0.51 for Neo. The music in Fig. 9 is medium-high pitch, medium high tempo and atonal, and could sound like a quite fearful chase. Finally *The Godfather* opening scene is estimated as valence 0.32, arousal 0.01 and dominance 0.68. The music suggested in Fig. 10 is low pitch, very low tempo and atonal. This supports the mood of the opening which is an unpleasant story told quietly in a dark office.

Because of the simplicity of the algorithms there will be an inevitable sameness to the tunes over time. However the tunes above could certainly be candidates to make a composer think about how to approach those scenes.

Although the system is designed as a creative tool rather than a creativity engine itself, it is interesting to consider the level of creativity of the algorithm. Looking at Section 5.2, this level is fairly low. The phrase generation is random with length fixed by the user (either six or eight notes, randomly selected in each example above). The theme generation is based on just five fixed rules to transform and concatenate phrases. The section generation is fixed by the user, for the above examples it was ABCA. However this lack of algorithmic creativity can be separated, to a degree, from how useful TRAC may be as a creative tool. What is clear is this: where there was nothing there is now something. The composer has been given ninety to two-hundred pages of narrative text they are unfamiliar with, which need to be turned into a similar number of minutes of music in some cases. Feeling emotionally overwhelmed by a musical task will not promote musical creativity. TRAC provides one tool for dealing with the “empty score” syndrome – a way to understand a new script a little more, and to start filling in some possible music features.



Figure 7 – First six lines of suggested music for the Alien revealing itself in a craft Ripley has used

to escape (*Alien* vPage 29) and *Alien* respectively.



Figure 8 – First six lines of suggested music for the Millennium Falcon's escape from the Death Star (*Star Wars* vPage 38).

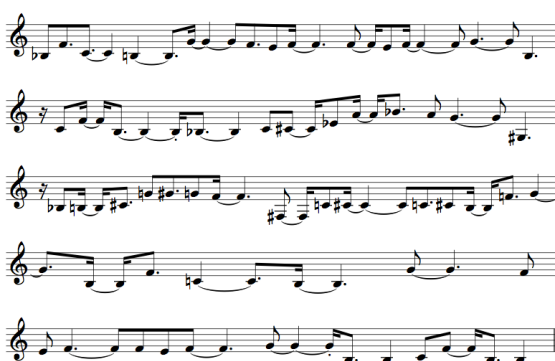


Figure 9 – Suggested music, based on Neo's affect towards the end of the film: Neo tries to escape from Agent Smith (*The Matrix* vPage 61).

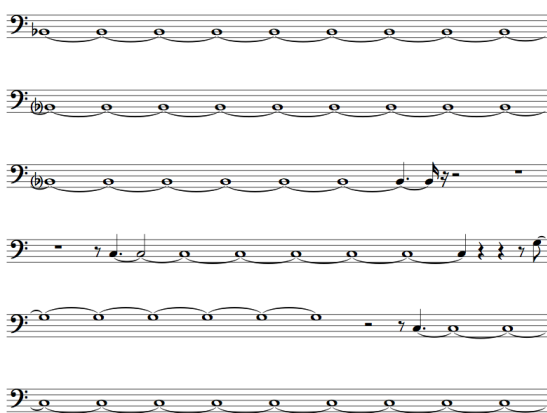


Figure 10 – First six lines of music suggested for the opening: a monologue at a wedding about a daughter being attacked (*The Godfather* vPage 0).

5 Conclusions and Future Work

Reading and absorbing scripts is a time consuming process for the feature sound-track composer. Developing an understanding of the script structure is an additional strain. In a bid to help address some of the pressures of film production, a system called *TRAC* (Textual Research for Affective Composition) has been developed. *TRAC* attempts to parse a film script and highlight structural elements. It also provides tools to generate short soundtrack sketches based on these affective elements. The process is very approximate, but for a composer under time-pressure, who is approaching a script for the first time, it can provide a valuable "in" for kick-starting and guiding some of their work process.

Examples have been given from four so-called "blockbuster" films: showing how the film's structure can be understood more rapidly through graphs of measures such as Scene Rate, Dialogue per vPage, and Character Activity. In addition an affective script analysis on a global or character-by-character basis can provide some limited insights. It also acts as a starting point from which to sketch soundtrack elements.

The current sound-track sketching system in *TRAC* is based on an extended random walk algorithm to generate short phrases. These phrases and their variations are then repeated and combined to create themes. The themes are combined into patterns such as ABA and ABCA, to create sections of music. The section is then transformed based on the affective data on a given vPage for a single

character or for all activity on the vPage. This transformed music is then presented to the composer as the example sketch. The composer can also hand-craft a protagonist motif or film motif, and instruct *TRAC* to incorporate these as phrases, themes, or sections into the generative process.

There are three clear areas for future work. One is formal testing of the affective composition system. The system is mostly based on previous research in affective algorithmic composition. However the precise combination and parameter ranges utilised have not been formally tested across a significant and independent population of listeners.

The second area for further work is the examination of the accuracy of the word-based affective estimates. Such an examination would be fairly complex. Film experts would markup twenty to thirty scripts by vPage – both term of protagonist emotion and the emotion of the whole vPage. These mark-ups would be compared to *TRAC*'s affective estimates.

A third area for further work would involve extending the underlying algorithmic composition tool. For example, a system based on higher-order Markov Models (Williams et al., 2015) could be trained using soundtracks from films of different genres. These models could be labelled by movie genre, thus allowing the composer to select the most effective for the type of script being examined. Another example is inspired by the recognition that the current system focuses on writing music which communicates a similar affect to that of the underlying text. However, there are many examples of sound-tracking where the music's affective content is at odds with that of the underlying scene. For example, slow minor key music has been used during tragic segments of high arousal battle scenes, emphasising the tragedy. Thus there is scope in *TRAC* to provide options for the user to control musical deviations from the estimated valence and/or arousal of a scene.

References

- Bellegarda, J. (2010). Emotion analysis using latent affective folding and embedding. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text* (pp. 1–9). Los Angeles, CA: Association for Computational Linguistics.
- Bettinson, H. (2013). Sounds of Hong Kong cinema. *Jump Cut: A Review of Contemporary Media*, (55).
- Bird, S., Klein, E. & Loper, E. (2009). *Natural language processing with Python*. Sebastopol, CA: O'Reilly Media, Inc.
- Boltz, M., (2004). The cognitive processing of film and musical soundtracks. *Memory and Cognition*, 32(7), 1194–1205.
- Ellis, R. & Simons, R. (2005). The impact of music on subjective and physiological indices of emotion while viewing films. *Psychomusicology*, 19(1), 15–40.
- Blumstein, D., Davitian, R. & Kaye, P. (2010). Do film soundtracks contain nonlinear analogues to influence emotion? *Biology Letters*. Retrieved from: <http://classic.rsbl.royalsocietypublishing.org/content/early/2010/05/20/rsbl.2010.0333.abstract> [Accessed: May 2, 2015].
- Bradley, M. & Lang, P. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Technical Report C-1. University of Florida: Center for Research in Psychophysiology.
- Clark, A., Fox, C. & Lappin, S., (2012). *The handbook of computational linguistics and natural language processing*. New York: Wiley.
- Cohen, A. (2013). Film music and the unfolding narrative. In Arbib, M.A. (Ed.), *Language, Music and the Brain* (pp. 173–201) (Strüngmann Forum Reports). Cambridge, MA: MIT Press
- Coppola, F. (1972). *The Godfather*. USA: Paramount Pictures.
- Davis, M., Hull, J., Young, R. & Warren, G. (1987). Emotional reactions to dramatic film stimuli: The influence of cognitive and emotional empathy. *Journal of Personality and Social Psychology*, 52(1), 126–133;.
- Doppler, J., Rubisch, J., Jaksche, M. & Raffaseder, H. (2011). RaPScoM: Towards composition strategies in a rapid score music prototyping framework. In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound* (pp. 8–14). Coimbra, Portugal: ACM.
- Eliashberg, J., Hui, S. & Zhang, Z. (2007). From story line to box office: A new approach for green-lighting movie scripts. *Management Science*, 53(6), 881–893.
- Fernandez, J. & Vico, F. (2013). AI methods in algorithmic composition: A comprehensive survey. *Journal of Artificial Intelligence Research*, 48, 513–582.
- Galician, M. & Bourdeau, P. (2008). The evolution of product placements in Hollywood cinema. *Journal of Promotion Management*, 10(1–2), 15–36.
- Gomes, J., Silva, F. & Chambel, T. (2013). Genetic soundtracks: Creative matching of audio to video. *arXiv preprint*, arXiv:1304.2671.
- Green, J. 2010. Understanding the score: Film music communicating to and influencing the audience. *Journal of Aesthetic Education*, 44(4), 81–94.
- Kulezic-Wilson, D. (2015). Music as model and metaphor In Kulezi-Wilson, D. (Ed.), *The musicality of narrative film* (pp. 18–33). London: Palgrave Macmillan.

- Le Groux, S. & Verschure, P. (2012). Subjective emotional responses to musical structure, expression and timbre features: A synthetic approach. In *9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)* (pp. 160–176). London: Queen Mary University of London.
- Hedemann, C., Sorensen, A. & Brown, A. (2008). Metascore: User interface design for generative film scoring. In *Proceedings of Sound: Space & The Australasian Computer Music Conference* (pp. 25–30). Sydney: ACMA.
- Hoyt, E., Ponto, K. & Roy, C. (2014). Visualizing and analyzing the Hollywood screenplay with ScripThreads. *Digital Humanities*, 8(4).
- Huang, C. & Lin, E. (2013). An emotion-based method to perform algorithmic composition. In *Proceedings of the 3rd International Conference on Music and Emotion (ICME3)*, Jyväskylä, Finland (pp. 244–247). Jyväskylä, Finland: Department of Music, University of Jyväskylä.
- Jewell, M., Nixon, M. & Prugel-Bennett, A. (2003). CBS: A concept-based sequencer for soundtrack composition. In *Proceedings of the IEEE Third International Conference on Web Delivering of Music*, Leeds, UK (pp. 105–108). IEEE.
- Jewell, M. (2007). *Motivated music: Automatic soundtrack generation for film*. PhD. Dissertation, University of Southampton, UK.
- Karlin, F. & Wright, R. (2004). *On the track: A guide to contemporary film scoring*. New York: Routledge.
- Kirke, A. & Miranda, E. (2011). Combining EEG frontal asymmetry studies with affective algorithmic composition and expressive performance models. In *Proceedings of International Computer Music Conference (ICMC 2011)*. Huddersfield: University of Huddersfield.
- Kirke, A. & Miranda, E. (2015). A multi-Agent emotional society whose melodies represent its emergent social hierarchy and are generated by agent communications. *Journal of Artificial Societies and Social Simulation*, 18(2), 16.
- Kirke, A. (2016a). http://cmr.soc.plymouth.ac.uk/alexiskirke/Example1_Kirke.aiff.
- Kirke, A. (2016b). http://cmr.soc.plymouth.ac.uk/alexiskirke/Example2_Kirke.aiff.
- Kirke, A. (2016c). http://cmr.soc.plymouth.ac.uk/alexiskirke/Example3_Kirke.aiff.
- Kirke, A. (2016d). http://cmr.soc.plymouth.ac.uk/alexiskirke/Example3_Kirke.aiff.
- Kundu, A., Das, D. & Bandyopadhyay, S. (2013). Scene boundary detection from movie dialogue: A genetic algorithm approach. *Polibits* 47, 55–60.
- Kuo, F., Shan, M. & Lee, S. (2013). Background music recommendation for video based on multimodal latent semantic analysis. In *Proceedings of the 2013 IEEE International Conference on Multimedia*, San Jose, CA (pp. 1–6). IEEE.
- Lin, G. & Walker, M. (2011). All the world's a stage: Learning character models from film. In *Proceedings of AIIDE 2011: AI and Interactive Digital Entertainment Conference* (pp. 46–52). Stanford, Palo Alto, CA: The AAAI Press.
- Lin, Y., Tsai, T., Hu, M., Cheng, W. & Wu, J. (2014). Semantic-based background music recommendation for home videos. In Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H. & O'Connor, N. (Eds.), *MultiMedia Modeling* (pp. 283–290). Heidelberg: Springer International Publishing.
- Livingstone, S., Muhlberger, R., Brown, A. & Thompson, W. (2010). Changing Musical Emotion: A computational rule system for modifying score and performance. *Computer Music Journal*, 34(1), 41–64.
- Lucas, G. (1977). *Star Wars*. USA: 20th Century Fox.
- Marsella, S., Gratch, J. & Petta, P. (2010). Computational models of emotion. In Scherer, K.R., Bänziger, T. & Roesch, E. (Eds.), *A Blueprint for Affective Computing: A Sourcebook and Manual* (pp. 21–46). New York: Oxford University Press.
- Miranda, E. (2001). *Composing music with computers*. Oxford: Focal Press.
- Munishkina, L., Parrish, J. & Walker, M. (2013). Fully-automatic interactive story design from film scripts. In Oyarzun, D., Peinado, F., Michael Young, R., Elizalde, A. & G. Méndez (Eds.), *Interactive Storytelling* (pp. 229–232). Heidelberg: Springer International Publishing.
- Murtagh, F. & Ganz, A. (2015). Pattern recognition in narrative: Tracking emotional expression in context. *Journal of Data Mining and Digital Humanities* (in press).
- Park, S., You, E. & Jung, J. (2012). Potential emotion word in movie dialog. In *Springer Lecture Notes in Electrical Engineering*, 120, 507–516.
- Sang, J. & Xu, C. (2010). Character-based movie summarization. In *Proceedings of the 2010 International Conference on Multimedia*, Firenze, Italy (pp. 855–858). New York: ACM.
- Scott, R. (1979). *Alien*. USA: 20th Century Fox.
- Sloboda, J. (1986). *The musical mind: The cognitive psychology of music*. Oxford: Oxford University Press.

- Seashore, C. (1938). *The psychology of music*. New York: Dover Publications.
- Stupar, A. (2013). *Soundtrack recommendation for images*. PhD Dissertation, Universität des Saarlandes.
- Tanrisever, B. (2001) *Opera functioning as narrative in films*. PhD Dissertation, Bilkent University.
- Vane, E. & Cowan, W. (2007). A Computer-aided soundtrack composition system designed for humans. In *Proceedings of the 2007 International Computer Music Conference, Copenhagen, Denmark* (pp. 192–199). San Francisco, CA: ICMA.
- Wachowski, L. & Wachowski, A. (1999). *The Matrix*. USA: Warner Bros.
- Walker, M., Grant, R., Sawyer, J., Lin, G., Wardrip-Fruin, N. & Buell, M. (2011). Perceived or not perceived: Film character models for expressive natural language generation. In *Interactive Storytelling: Lecture Notes in Computer Science 7069* (pp. 109–121). Heidelberg: Springer International Publishing.
- Warriner, A., Kuperman, V. & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research methods*, 45(4), 1191–1207.
- Williams, D., Kirke, A., Eaton, J., Miranda, E., Roesch, E., Daly, I. & Nasuto, S. (2014). Investigating affect in algorithmic composition systems. *Psychology of Music* (in press).
- Williams, D., Kirke, A., Eaton, J., Miranda, E., Daly, I., Hallowell, J., Roesch, E., Hwang, F. & Nasuto, S. (2015). Dynamic game soundtrack generation in response to a continuously varying emotional trajectory. In *Proceedings of 56th Audio Engineering Society Conference*. London: Audio Engineering Society.

Notes

There are no notes.

Author Contacts

Alexis Kirke

Alexis.kirke@plymouth.ac.uk

Plymouth University Interdisciplinary Centre for Computer Music Research

Plymouth University

Plymouth, PL4 8AA

United Kingdom

Eduardo Miranda

Eduardo.miranda@plymouth.ac.uk

Plymouth University Interdisciplinary Centre for Computer Music Research

Plymouth University

Plymouth, PL4 8AA

United Kingdom

Back to top

© 2016 - Journal of Creative Music Systems | ISSN 2399-7656

University of
HUDDERSFIELD
University of Huddersfield Press